CLEMSON

UNIVERSITY

ASPECTS OF STATISTICAL FILTER THEORY

Department of Mathematics

AUGUST 1965

Clemson University

Department of Mathematics

ASPECTS OF
STATISTICAL FILTER THEORY

August 1965

A Report on Contract NAS 8-11259

Advanced Studies Office

Astrionics Laboratory

Marshall Space Flight Center

National Aeronautics and Space Administration

Huntsville, Alabama

CLEMSON UNIVERSITY
INVESTIGATORS

C. V. Aucoin

J. W. Kenelly, Jr.

S. M. Lukawecki

B. J. Prochaska

Table of Contents

Illustrations

List of Symbols by Sections

History and Introduction

t	time in general; present time
$x(t)$	state n-vector
$F(t)$	n x n matrix
$D(t)$	n x m matrix
$u(t)$	input m-vector
$y(t)$	output p-vector
$M(t)$	n x p matrix
$\Phi(t+1; t)$	n x n transition matrix
$\Delta(t)$	n x m matrix
ε	error in estimation
$L(\varepsilon)$	loss function

Problems and Solutions

$u(t)$	independent random vector process
$Q(t)$	covariance matrix of $u(t)$
$\hat{y}(t_n)$	estimate of $y(t_n)$

Kalman's Derivation

$\mathcal{Y}(t)$	linear manifold
$\mathcal{Z}(t)$	linear manifold
$\bar{x}(t+1\|t)$	orthogonal projection of $x(t+1)$ on $\mathcal{Y}(t)$
$\tilde{x}(t+1\|t)$	component of $x(t+1)$ orthogonal to $\mathcal{Y}(t)$; error in optimal estimate
$x^*(t+1\|t)$	optimal estimate of $x(t+1)$

$\tilde{y}(t\|t-1)$	component of y(t) orthogonal to $\mathcal{Y}(t-1)$
$\hat{E}[x(t+1)\|\ \mathcal{Y}(t)]$	also optimal estimate of x(t+1); orthogonal projection of x(t+1) on $\mathcal{Y}(t)$
$\hat{E}[x(t+1)\|\ \mathcal{Z}(t)]$	orthogonal projection of x(t+1) on $\mathcal{Z}(t)$
$x^*(t+1\|t)$	best estimate of x(t+1)
$\Delta^*(t)$	n x p matrix
$\Phi^*(t+1;\ t)$	n x n matrix
$\tilde{x}(t+1\|t)$	error in estimation of x(t+1)
$P^*(t)$	covariance matrix of the error in estimation at time t
t_o	time at which observations start
$\hat{x}(t_o)$	estimate of $x(t_o)$
E	expectation
$\tilde{x}(t_o)$	error in estimate of $x(t_o)$
$\hat{x}^*(t_n)$	updated estimate of $x(t_n)$
$\hat{x}(t_n)$	estimate of $x(t_n)$
$P(t_n)$	covariance matrix of the error in the estimate $\hat{x}(t_n)$
$\bar{y}(t\|t-1)$	orthogonal projection on $\mathcal{Y}(t-1)$

Example

A(t)	4 x 4 matrix
$S(t_2;\ t_1)$	series expansion of an element of the Φ matrix
$R(t_2;\ t_1)$	series expansion of an element of the Φ matrix
$S_k(t_2;\ t_1)$	k terms of $S(t_2;\ t_1)$
$R_k(t_2;\ t_1)$	k terms of $R(t_2;\ t_1)$
$\Phi_k(t_2;\ t_1)$	approximation of $\Phi(t_2;\ t_1)$

Abstracts of Selected Reports

J.P.L.[1]

Λ_x	expectation of xx'
Λ_y	expectation of yy'
Λ_{yx}	expectation of yx'
$\hat{x}(t_{n+1})$	"best" estimate of the error in estimation
$\hat{y}$	"best" linear estimate of y
Q	covariance matrix of the error in the estimate of $\hat{y}$
Q_m	covariance matrix with minimum trace

Autonetics [5]

a_{n-1}	control filter
$C(\Delta\tilde{y}_n)$	covariance matrix on observation errors
$C(\Delta\hat{x}_n')$	covariance matrix on dispersion of predicted estimate
M_n	generalized matrix relating current position-velocity deviation to observables
u_n	generalized trajectory control (velocity-to-be-gained)
x_n	position-velocity deviation state vector at beginning of the n^{th} cycle
$\hat{x}_n$	optimal estimate of deviation state
x_n^A	actual deviation state
y_n	generalized observation residuals vector for the navigation-guidance cycle
y_n^A	actual observation residuals vector
$\Delta\tilde{y}_n$	generalized error in observables

$\hat{y}'_n$	optimal prediction of observation residuals vector
$\tilde{y}_n$	measured or computed observation residuals vector
$\hat{y}_n$	optimal estimate of observation residuals vector
$\Delta_{n+1,\,n}$	matrix relating present velocity correction at n to predicted state at n+1
$\Phi_{n+1,\,n}$	transition matrix
ω_n^*	data process filter for optical estimate of observables
ω_n	data process filter for optical estimate of deviation state
$C(\Delta\hat{x}'_n)$	covariance matrix on dispersion of predicted estimate
$C(\Delta\tilde{y}_n)$	covariance matrix on observation errors

Philco [10]

$X(t)$	observable variable
y	vector of the coefficient of a polynomial
$\hat{y}$	least squares choice for y
P_x	covariance matrix of the observations made on the variable X(t)
$P_{\hat{y}}$	covariance matrix of $\hat{y}$
$\hat{x}$	estimate of x
H	m x n matrix
q	error in measurement of y

Q	covariance matrix of the error q
$\hat{x}_n$	new estimate of x
$p(x\|y, \hat{x})$	conditional probability density function
P_n	covariance matrix of the new estimate

Bissett-Berman [12]

$y(t)$	variable
$y_m(t)$	measurement of the variable y(t)
$p(a, b)$	likelihood function of parameters a and b
$\hat{a}$	estimate of parameter a
$\hat{b}$	estimate of parameter b
$\mathcal{L}$	function to be minimized
$p(a)$	likelihood function of parameter a

Summary

This report is intended to introduce and partially survey aspects of statistical filter theory and its relationship to space navigation problems. Space navigation dynamics are generally represented by non-linear systems. There are methods by which certain non-linear systems may be approximated by linear systems, and thus, the results applicable to linear systems may be employed. Attention will be directed primarily to the details of multivariable linear filter theory and reference will be given to the linearization problems.

This material may be of interest to persons of varied backgrounds and interests. In general, the interest of an individual may be in a brief survey. This individual is suggested to follow outline one as given below. In the event of interest in a more detailed investigation, outline two may be followed. A brief survey requires a minimun background in differential equations, matrix theory, elementary mathematical statistics, and navigational terminology. Whereas, the detail investigation is predicated on knowledge of linear algebra, systems theory, mathematical statistics, probability theory, and navigational problems.

The authors acknowledge the assistance of Mrs. Ann Elliott, the typist.

Outline 1

Section 1
↓
Section 2A
↓
Section 3
↓
Literature

[1] [1], [2]

Outline 2

Section 1
↓
Section 2
↓
Section 3
↓
Section 4
↓
Literature

[2] , [3], [4], [5]

[1]Numbers in brackets designate references in the Bibliography

Abstract

The Mathematics Department of Clemson University under NASA contract NAS8 - 11259 investigated aspects of statistical filter theory. The investigation included abstracting selected reports pertaining to R. E. Kalman's statistical filtering technique and developing a mathematical model to illustrate and study applications of statistical filtering. This investigation brought out certain areas in which further study may be beneficial.

1. History and Introduction

Classically, the problem of fitting a curve to a set of data points was attacked by the method of least squares. The underlying concept of the least squares method is that the fitted curve is selected in such a manner that the sum of the squares of the deviations of the data from the curve is minimized. This technique does not take into account the fact that some of the data may be more accurate than the other data. But this method is still very useful in many experimental problems.

The least square method may be modified to incorporate statistical information about the accuracy of the data, i.e., the data is weighed with regard to the associated statistics. Usually a curve is fitted to the data with each information particle incorporated via considerations regarding the related variance and covariances, along with the square of its deviation from the fitted curve. It differs from the usual least squares technique in this evaluation criteria, thus, the name weighted least squares is usually assigned.

In the general situation of data filtering, the information inputs may be treated as random variables or as deterministic values contaminated with random noise. That is, instead of fitting a curve, the basic signal of a known form is assumed to be corrupted with noise and it is desired to extract the signal by a linear system with minimum error in accordance with a chosen loss criteria. In his pioneering work Weiner [6] showed that the general problem of detecting a signal in the presence of random noise lead to the so called Weiner-Hopf equations. In particular,

the solution consists of a systematic method for determining the weighing function of a linear system that is optimum in the sense that the mean-square error between the actual output and the desired output is minimized, i.e., the statistical information of the inputs give the weighing function that yields an output differing with minimum mean-square error from the output of the noise free system. Various extensions of this theory have been developed, but the application of the solution is usually handicapped by many practical considerations.

An alternate approach to the linear filtering problem was formulated by Kalman [3] utilizing state space methods and linear dynamic systems. This new approach is especially significant. First of all in the previous theoretical structures the complete state vector is assumed to be available for use in the calculation, but this seldom is the practical case and measurements of only a few of the state variables may be available. The Kalman approach is able to process incomplete state variable data. Secondly, the Kalman approach adapts quite readily to computational routines. The Kalman approach is the main consideraton of this report and it will now be treated in considerable detail.

To adequately define the problem and solution, some basic terminology will be introduced. The notation will be identical with Kalman[3] except where indicated. The equations will be referred to via the sectionally sequential numbers in parentheses and the numbers as given by Kalman [3] will be enclosed in braces for cross reference purposes.

A system is a mathematical abstraction that is devised to serve as a model for a dynamic phenomena. The inputs represent, in the form of a set of time functions or sequences, the external forces that are acting upon the dynamic phenomena. The outputs represent the measures of the directly observable behaviour of the phenomena. The state of a dynamic system at time t is a set of numbers, called state variables, such that the knowledge of the numbers and the input will, with the equations describing the dynamics, provide the future state and output of the system. The set of all time values for which the inputs, the states, and the outputs are defined is termed the time space. If the time space is continuous, the system is known as a continuous time system. If the input and the state vectors are defined only for discrete instants of time t_k, where k ranges over the integers, the time space is discrete and the system is referred to as a discrete-time system.

A linear continuous time dynamic system may be described in general by the vector differential equation

$$dx(t)/dt = F(t)x(t) + D(t)u(t) \qquad (1.1a)$$ {12, [3]}

and

$$y(t) = M(t)x(t), \qquad (1.1b)$$ {12, [3]}

where $x(t)$[1] is a state n-vector, $u(t)$ is an input m-vector ($m \leq n$)

---

[1]Lower case letters will be used to indicate column vectors and upper case letters will be used to indicate matrices.

F(t) and D(t) are n x n and n x m matrices respectively. Finally, y(t) is an output p-vector of the system; and M(t) is an n x p matrix ($p \leq n$).

A linear discrete time dynamic system may be described by the vector difference equation

$$x(t + 1) = \Phi(t + 1; t)x(t) + \Delta(t)u(t) \quad (1.2a)$$ {14, [3]}

and

$$y(t) = M(t)x(t). \quad t = 0, 1, 2, \ldots \quad (1.2b)$$ {14, [3]}

Here $\Phi(t + 1; t)$ is the transition matrix of the equation (1.2a) and the matrix $\Delta(t)$ is of dimension of n x m. This is a particular formulation of a discrete system in which the observations are measured at integer values. In general the discrete system does not have to be integer valued or equally spaced. It is noted at this point that Kalman [7] gives a precise mathematical definition of a dynamical system and he gives conditions for when equations (1.1a) and (1.1b) represent a linear continuous time dynamical system.

In practical considerations of noise as related to discrete linear systems, the general structure may be investigated by three different associated methods of approach. Kalman [3] formulates his system with the linearly related output and random additive excitations in the linear system, i.e., the relationships are

$$x(t + 1) = \Phi(t + 1; t)x(t) + u(t), \quad (1.3)$$

$$y(t) = M(t)\, x(t).$$

Solloway [1] in his system formulation has additive random errors in his observations and no additive random excitation in the linear system, i.e., the relationships are

$$x(t_{n+1}) = \Phi(t_{n+1};\, t_n)\, x(t_n),$$
$$y(t_n) = M(t_n)\, x(t_n) + u(t_n). \tag{1.4}$$

A more general approach is given by Joyce [8] and Leibelt and Bergson [4] in which there is additive randomness in the linear system and observations, i.e., the relationships are given by

$$x(t_{n+1}) = \Phi(t_{n+1};\, t_n)x(t_n), \tag{1.5}$$
$$y(t_{n+1}) = M(t_n)x(t_n) + v(t_n).$$

The basic mathematical and statistical approaches in deriving the optimal filters also varied. Kalman derived his filter for an optimality criterion of minimum mean square error, using an orthogonal projection technique, while Solloway derived his results from the viewpoint of a minimum variance estimate.

Optimality criteria are basically dependent upon the concept of loss functions. Intuitively, a loss function assigns a penality to the error made in estimation. A loss function L generally is a function of the difference ε between the quantity to be estimated and the estimator. Formally, a loss function should be characterized by the following properties;

$$\text{(i)}\quad L(0) = 0,$$
$$\text{(ii)}\quad L(\varepsilon_1) \leq L(\varepsilon_2) \text{ for } 0 \leq \varepsilon_1 \leq \varepsilon_2, \tag{1.6}$$
$$\text{(iii)}\quad L(\varepsilon) = L(-\varepsilon).$$

For example, a popular loss function is $L(\varepsilon) = \varepsilon^2$.

In the employment of a loss function, $L(\varepsilon)$, a particular problem could be to derive a solution that would minimize the expected loss, i.e., the expectation of the loss function. Such a solution would be termed an optimal solution. This concept will be used and further explained in the subsequent derivation.

2. The Filters

A. Problems and Solutions.

Kalman's statistical filter is a particular formulation of a solution to the Weiner problem. Kalman [3] expresses the underlying situation as a dynamic model

$$x(t + 1) = \Phi(t + 1; t)\, x(t) + u(t), \qquad (2.1)\ \{16, [3]\}$$

$$y(t) = M(t)\, x(t), \qquad (2.2)\ \{17, [3]\}$$

where $u(t)$ is an independent normally distributed random process of n-vectors with zero mean and covariance $Q(t)$, $x(t)$ is an n-vector, $y(t)$ is a p-vector ($p \leq n$), and $\Phi(t + 1; t)$, $M(t)$ are $n \times n$ and $p \times n$ respectively with elements as non-random functions of time. The problem is to find an estimate, $x^*(t + 1 \mid t)$ of $x(t + 1)$ which minimizes the expected loss; given the observed values $y(t_0), \ldots, y(t)$ that are assumed to be linearly independent.

The loss function is assumed to satisfy the formal characterization of a general loss function as given in (1.6). The assumed normal distribution hypothesis implies that the optimal estimate would be the same as if the optimal estimate was restricted to a linear function of the observed random variables and a loss function $L(\varepsilon) = \varepsilon^2$.

Kalman's solution is an iterative scheme. The estimate $x^*(t + 1 \mid t)$ of $x(t + 1)$ minimizes the expected loss and it is

based on the observed outputs, $y(t_0)$, $y(t_1)$, ..., $y(t)$. This estimate satisfies a dynamic system

$$x^*(t+1 \mid t) = \Phi^*(t+1;\, t)x^*(t \mid t-1) + \Delta^*(t)y(t), \tag{2.3}$$

{21, [3]}

where for $t \geq t_0$

$$\Delta^*(t) = \Phi(t+1;\, t)\, P^*(t)M'(t)\,[M(t)P^*(t)M'(t)]^{-1}, \tag{2.4}$$

{28, [3]}

$$\Phi^*(t+1;\, t) = \Phi(t+1;\, t) - \Delta^*(t)M(t). \tag{2.5}$$

{29, [3]}

The error in estimation $\hat{x}(t+1 \mid t)$ satisfies the system

$$\tilde{x}(t+1 \mid t) = \Phi^*(t+1;\, t)\,\tilde{x}(t \mid t-1) + u(t). \tag{2.6}$$

{23, [3]}

The covariance matrix of the error in estimation is $P^*(t)$ and it satisfies the recursion formula

$$P^*(t+1) = \Phi^*(t+1;\, t)\, P^*(t)\Phi'^*(t+1;\, t) + Q(t). \tag{2.7}$$

{30, [3]}

A basic underlying assumption is that an estimate at time t_0, $\hat{x}(t_0)$, and the associated covariance matrix of the error in estimate is known. This is necessary to initiate the iterative process in Kalman's solution. Kalman also assumes that the expected value of the error, in the original estimate, $\tilde{x}(t_0)$, is zero, i.e., $E[\tilde{x}(t_0)] = 0$. The covariance matrix of the error in the original estimate, is $P^*(t_0) = E[\tilde{x}(t_0)\tilde{x}'(t_0)]$.

---

[1]The prime notation on a matrix is used to indicate the transpose.

Solloway's model for the dynamic system is given by

$$x(t_{n+1}) = \Phi(t_{n+1};\, t_n)\, x(t_n), \tag{2.8a}$$

$$y(t_n) = M(t_n)\, x(t_n) + u(t_n), \tag{2.8b}$$

where $u(t_n)$ is an independent random process of p-vectors with zero mean and known covariance matrix, $Q(t_n)$, i.e.,

$$Q(t_n) = E\,[\,u(t_n)\, u'(t_n)],$$

$$E\,[u(t_n)\, u'(t_n)\,] = 0,\ n \neq m$$

$$E[u(t_n)] = 0 \text{ for all } n.$$

The problem is then: to find an updated estimate $\hat{x}^*(t_n)$ of the state $x(t_n)$ by using the observation $y(t_n)$ and an estimate $\hat{x}(t_n)$ based on the observations $y(t_o)$, $y(t_1)$, ..., $y(t_{n-1})$. The covariance matrix $P(t_n)$ of the error in estimate $\hat{x}(t_n)$ is assumed to be known. The criteria used in establishing the updated estimate $\hat{x}^*(t_n)$ is to minimize the trace of the covariance matrix $E[\{x(t_n) - \hat{x}^*(t_n)\}\{x(t_n) - \hat{x}^*(t_n)\}']$.

The solution of Solloway's model can be represented by

$$\hat{x}^*(t_n) = \hat{x}(t_n) + P(t_n)M'(t_n)(M(t_n)P(t_n)M'(t_n)+Q(t_n))^{-1}(y(t_n)-\hat{y}(t_n)) \tag{2.9}$$

where

$$\hat{y}(t_n) = M(t_n)\hat{x}(t_n), \tag{2.10}$$

$$\hat{x}(t_n) = \Phi(t_n;\, t_{n-1})\, \hat{x}^*(t_{n-1}), \tag{2.11}$$

$$P(t_n) = \Phi(t_n;\, t_{n-1})\, P^*(t_{n-1})\Phi'(t_n;\, t_{n-1}). \tag{2.12a}$$

The covariance matrix, $P^*(t_n)$, of the error in the estimate is

$$P^*(t_n) = P(t_n) - P(t_n)M'(t_n)(M(t_n)P(t_n)M'(t_n) + Q(t_n))^{-1}M(t_n)P(t_n). \quad (2.12b)$$

An initial estimate $\hat{x}(t_o)$ and the covariance matrix $P(t_o)$ of the error in the estimate is required to initiate the iterative scheme.

A computer orientated flow chart of the Solloway solution is illustrated in Figure 1. Notice that the step-wise estimation scheme requires knowledge of the linear dynamic system. This involves knowing the transition matrix, the mapping matrix and the covariance matrix of the additive random errors in the observations. In Figure 1, the observation schedule consists of the time t_o, t_1, ..., t_n, ... at which the observations will be made.

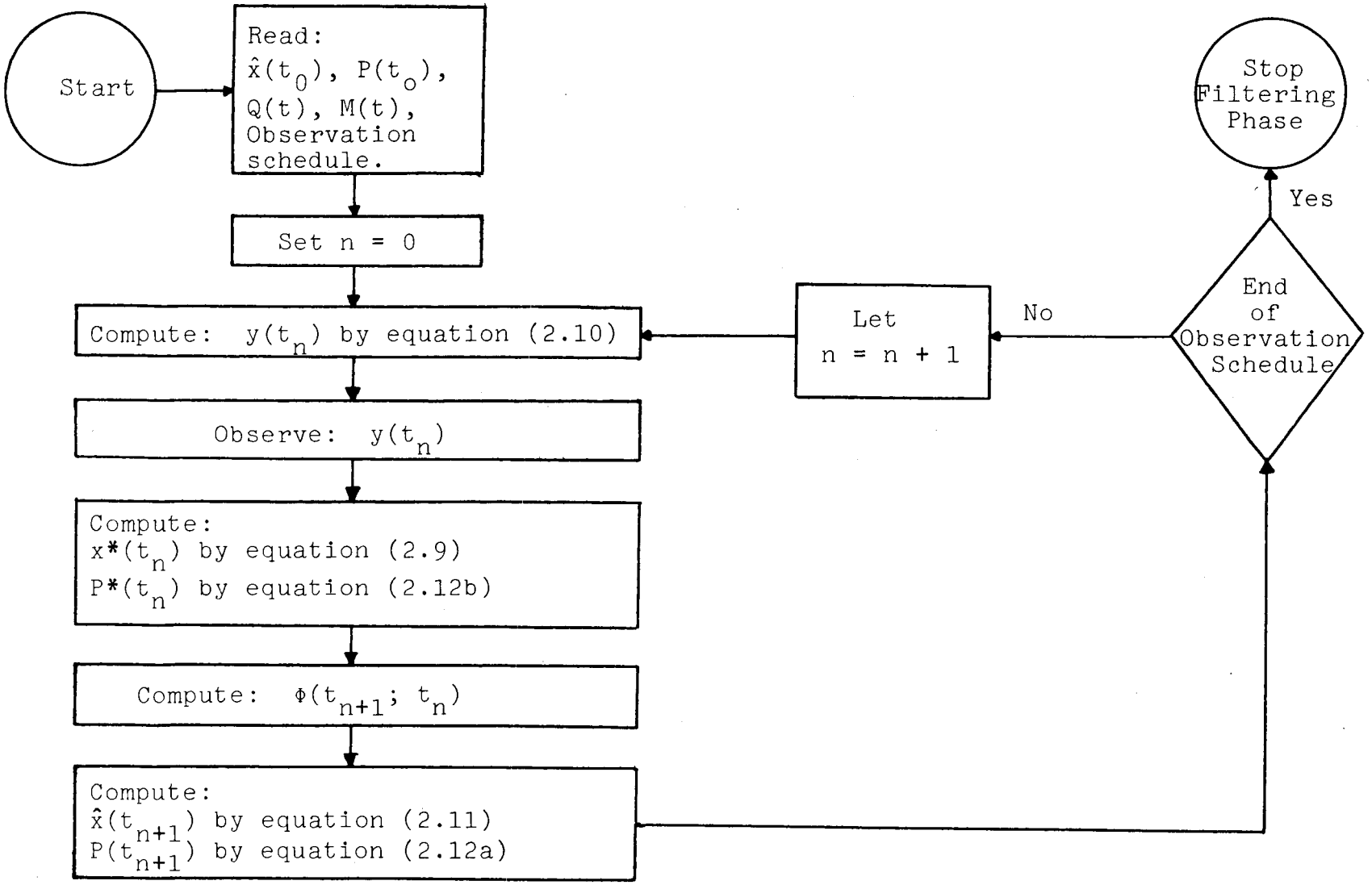

Figure 1. Flow Chart for the Solution of the Model Given in Equation (2.8)

B. Kalman's Derivation

There are notation conventions and definitions necessary for the derivation of Kalman's filter. The derivation is given in the framework of linear spaces generated by random variables. The observed random variables are treated within the linear manifolds that they generate, and the optimal estimates are expressed as orthogonal projections on the manifolds.

The linear manifold generated by the random variables $y(t_0), \ldots, y(t)$ is denoted by $\mathcal{Y}(t)$. When the observable $y(t)$ is a p-vector, the linear manifold $\mathcal{Y}(t)$ is the set of all linear combinations of all component of the vectors $y(t_0), \ldots, y(t)$, i.e.,

$$\sum_{i=t_0}^{t} \sum_{j=1}^{p} a_{ij} y_j(i)$$

where $y_j(i)$ is the j^{th} component of $y(i)$.

Let U and V be linear manifolds such that U is contained in V. A vector x of V is said to be orthogonal to the linear manifold U if and only if x is orthogonal to every vector of U. Furthermore, any vector w of V may be expressed uniquely as y + z with y an element of U and z orthogonal to U. For the vector w, the y satisfying the equation w = y + z is termed the projection of w on U.

The orthogonal projection of the random variable $x(t+1)$ on $\mathcal{Y}(t)$ is denoted as $\bar{x}(t+1 \mid t)$ and the component of $x(t+1)$ orthogonal to $\mathcal{Y}(t)$ is denoted as $\tilde{x}(t+1 \mid t)$. When $x(t+1)$ is an n-vector the projection of $x(t+1)$ on the mani-

fold $\mathcal{Y}(t)$ is considered to be a component-wise projection, i.e., the i^{th} component of $\bar{x}(t+1 \mid t)$ is the projection of the i^{th} component of $x(t+1)$ onto $\mathcal{Y}(t)$. The vector $\tilde{y}(t \mid t-1)$ which is the component of $y(t)$ orthogonal to $\mathcal{Y}(t-1)$ generates the linear manifold $\mathcal{Z}(t)$. Given $\mathcal{Y}(t)$ the optimal estimate of $x(t+1)$ is denoted by $x^*(t+1 \mid t)$ and the error in this optimal estimate is denoted by $\tilde{x}(t+1 \mid t)$.

The derivation of Kalman's version of the Wiener problem depends on a well known theorem [9] from probability theory. The theorem is stated without proof: Let $\{x(t)\}$, $\{y(t)\}$ be normally distributed random processes with zero mean, i.e., $E[x(t)] = 0$, $E[y(t)] = 0$ for all t. Then the optimal estimate, $x^*(t+1|t)$, of $x(t+1)$ given the observations $y(t_o), \ldots, y(t)$, is the orthogonal projection, $\bar{x}(t+1|t)$, of $x(t+1)$ on $\mathcal{Y}(t)$.

An alternate notation for $x^*(t+1 \mid t)$ which illustrates the dependency of the optimal estimate on the linear manifold $\mathcal{Y}(t)$ is $\hat{E}[x(t+1) \mid \mathcal{Y}(t)]$.

Assume $\mathcal{Y}(t-1)$ is known and the random variable, $y(t)$, at time t is observed. The orthogonal component $\tilde{y}(t \mid t-1)$ will generate the linear manifold $\mathcal{Z}(t)$. The linear manifold $\mathcal{Y}(t)$ is the same as the one obtained by the direct sum of $\mathcal{Y}(t-1)$ and $\mathcal{Z}(t)$. This is illustrated in Figure 2. By the definition of $\mathcal{Z}(t)$ every vector in $\mathcal{Z}(t)$ is orthogonal to every vector in $\mathcal{Y}(t-1)$.

Now the linear manifold $\mathcal{Y}(t)$ is the direct sum of $\mathcal{Y}(t-1)$ and $\mathcal{Z}(t)$ and thus,

$$x^*(t+1 \mid t) = \hat{E}[\, x(t+1) \mid \mathcal{Y}(t)\,]$$
$$= \hat{E}[x(t+1) \mid \mathcal{Y}(t-1)\,] + \hat{E}[\, x(t+1) \mid \mathcal{Z}(t)\,]. \quad (2.13)$$

This means that the orthogonal projection of $x(t+1)$ on $\mathcal{Y}(t)$ can be expressed as the sum of the orthogonal projection $x(t+1)$ on $\mathcal{Y}(t-1)$ plus the orthogonal projection of $x(t+1)$ on $\mathcal{Z}(t)$. This is illustrated in Figure 3.

Now using the inductive assumption that $x^*(t \mid t-1)$ is known, equation (2.13) can be written as

$$x^*(t+1 \mid t) = \hat{E}[\, \Phi(t+1;\, t)x(t) + u(t) \mid \mathcal{Y}(t-1)\,]$$
$$+\hat{E}[\, x(t+1) \mid \mathcal{Z}(t)]$$
$$= \Phi(t+1;\, t)\, \hat{E}[\, x(t) \mid \mathcal{Y}(t-1)\,]$$
$$\hat{E}[u(t) \mid \mathcal{Y}(t-1)\,] + \hat{E}[x(t+1) \mid \mathcal{Z}(t)\,] \quad (2.14)$$
$$\{18, [3]\}$$
$$= \Phi(t+1;\, t)\, x^*(t \mid t-1) + \hat{E}[\, u(t) \mid \mathcal{Y}(t-1)]$$
$$+ \hat{E}[\, x(t+1) \mid \mathcal{Z}(t)\,].$$

This results follows by noting that an orthogonal projection of a linear combination of two vectors is the same linear combination of the projection of the individual vectors.

Since the input is an independent random process, the vector $u(t)$ is independent of all the previous noise vectors. In particular, $u(t)$ is independent of $u(t-2)$, $u(t-3)$, ..., and therefore, by equations (2.1) and (2.2) $u(t-2)$ affects $x(t-1)$ which in turn affects $y(t-1)$. But $u(t)$ is independent of $u(t-2)$ and consequently, $y(t-1)$ is independent of $u(t)$. In a like

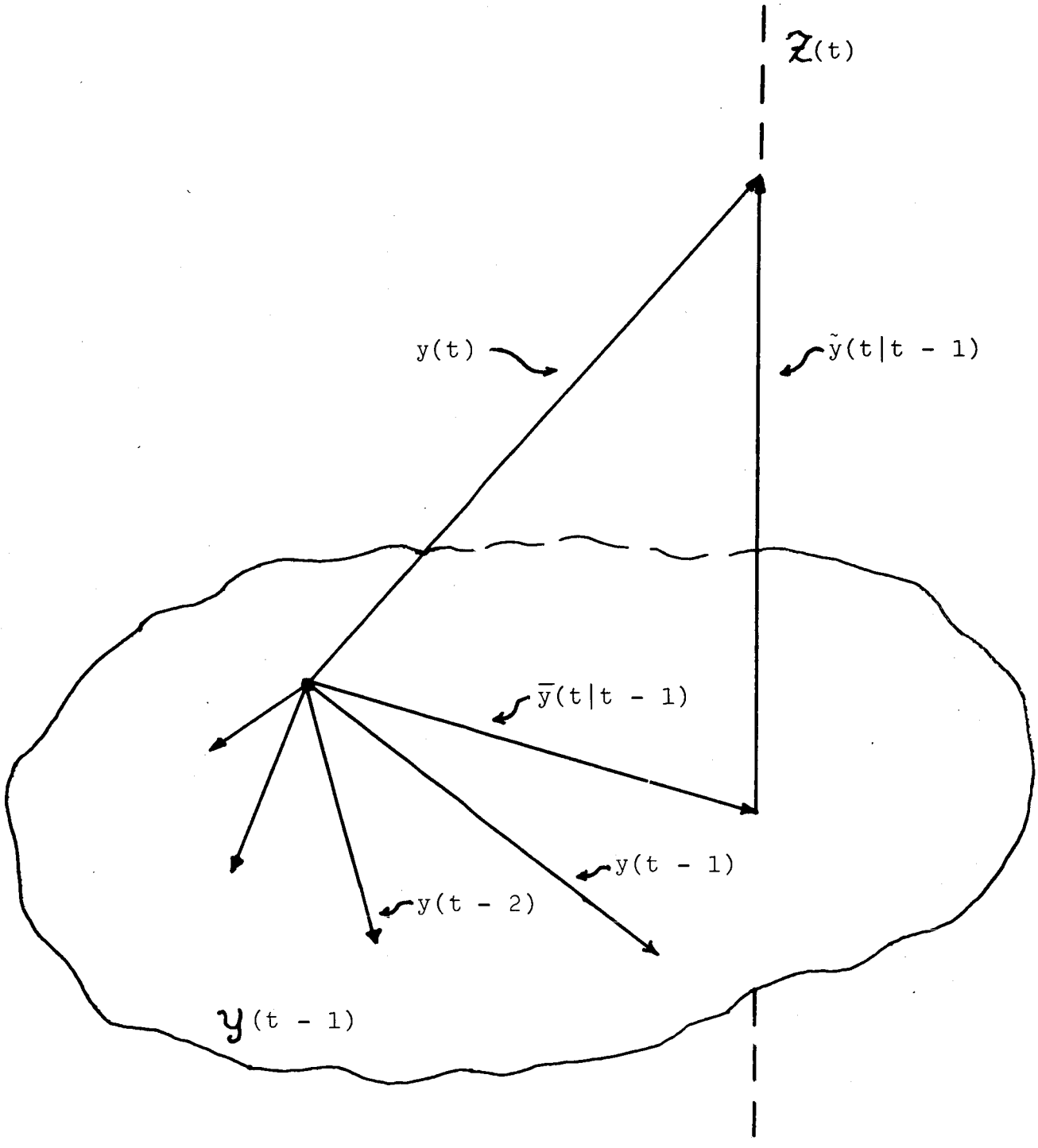

Figure 2

Schematic Representation of $\mathcal{Y}(t-1)$ and $\mathcal{Z}(t)$

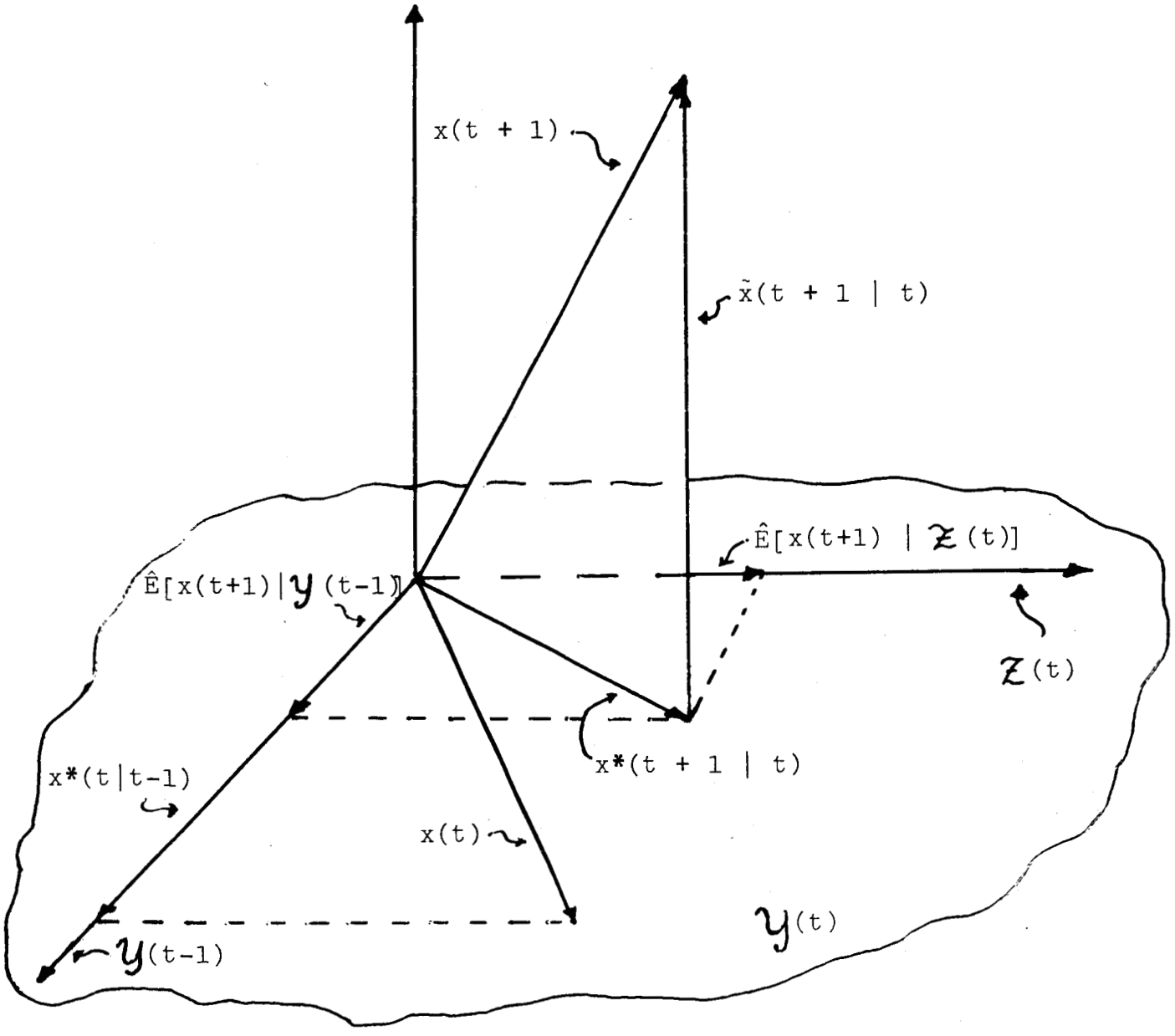

Figure 3

Schematic Representation of Equation (2.13)

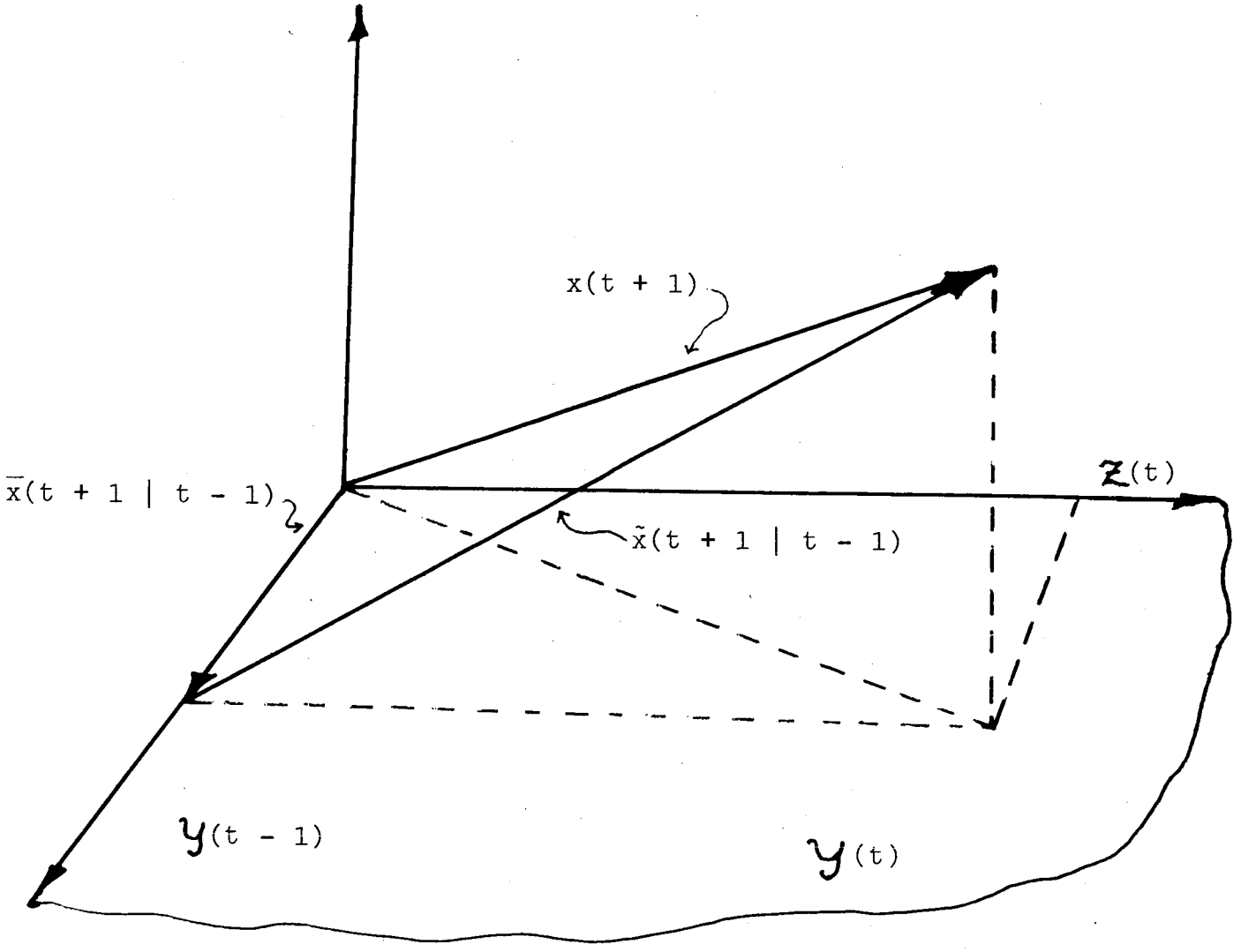

Figure 4

Schematic Representation of x(t + 1)

manner, u(t) is independent of y(t - 2), y(t - 3), ... As noted y(t - 1), y(t - 2), ..., $y(t_o)$ generates the linear manifold $\mathcal{Y}$ (t - 1). Therefore, u(t) is independent of $\mathcal{Y}$ (t - 1). Since E[u(t)] = 0, for all t and since u(t) is independent of $\mathcal{Y}$ (t - 1) u(t) is orthogonal to the linear manifold $\mathcal{Y}$(t - 1), i.e., each component of u(t) is orthogonal to $\mathcal{Y}$ (t - 1). Therefore,

$$\hat{E}[\, u(t) \mid \mathcal{Y}\ (t - 1)\,] = 0 \qquad (2.15)$$

Since $\hat{E}[\, x(t + 1) \mid \mathcal{Z}\ (t)\,]$ is the orthogonal projection of x(t + 1) on $\mathcal{Z}$(t) which is generated by $\tilde{y}(t \mid t - 1)$ the projection must be a linear operation on $\tilde{y}(t \mid t - 1)$, i.e.,

$$\hat{E}[\, x(t + 1) \mid \mathcal{Z}\ (t)\,] = \Delta^*(t)\tilde{y}(t \mid t - 1), \qquad (2.16)\{\ 19, [\,3\,]\}$$

where $\Delta^*(t)$ is an n x p matrix.

The representation of y(t) is

$$y(t) = \bar{y}(t \mid t - 1) + \tilde{y}(t \mid t - 1)$$

where $\bar{y}(t \mid t - 1)$ is a unique element of $\mathcal{Y}$(t - 1).

Express x(t) as $x^*(t \mid t - 1) + [\, x(t) - x^*(t \mid t - 1)\,]$, and use equation (2.2) to obtain

$$\bar{y}(t|t-1) + \tilde{y}(t|t-1) = M(t)x^*(t|t-1) + M(t)\,[\, x(t) - x^*(t|t-1)\,]$$

$$= M(t)x^*(t|t-1) + M(t)\,\tilde{x}(t|t-1). \qquad (2.17)$$

Consequently, $\bar{y}(t|t-1) = M(t)x^*(t|t-1)$ since $M(t)x^*(t|t-1)$ is the component of M(t)x(t) in $\mathcal{Y}$(t-1).

Hence, $\tilde{y}(t|t-1) = y(t) - \bar{y}(t|t-1)$

$$= y(t) - M(t)x^*(t|t-1). \qquad (2.18)\ \{20,\ 3\ \}$$

By substituting equations (2.15), (2.16), and (2.18) into equation (2.14)

$$x^*(t+1 \mid t) = \Phi(t+1;\, t)\, x^*(t \mid t-1)$$
$$+\Delta^*(t)[y(t) - M(t)x^*(t \mid t-1)]$$
$$= \Phi^*(t+1;\, t)x^*(t \mid t-1) + \Delta^*(t)y(t)$$
(2.19) {21, [3]}

where

$$\Phi^*(t+1;\, t) = \Phi(t+1;\, t) - \Delta^*(t)M(t). \qquad (2.20)\ \{22, [\ 3\ ]\}$$

Note that equation (2.19) represents a linear dynamic system.

A linear dynamic system also governs the estimation error, i.e.,

$$\tilde{x}(t+1 \mid t) = x(t+1 \mid t) - x^*(t+1 \mid t)$$
$$= \Phi(t+1;\, t)x(t) + u(t) - \Phi^*(t+1;\, t)x^*(t \mid t-1)$$
$$-\Delta^*(t)M(t)x(t)$$
$$=[\Phi(t+1;\, t) - \Delta^*(t)M(t)]x(t)$$
$$-\Phi^*(t+1;\, t)\, x^*(t \mid t-1) + u(t)$$
$$= \Phi^*(t+1;\, t)x(t) - \Phi^*(t+1;\, t)\, x^*(t \mid t-1)$$
$$+ u(t)$$
$$= \Phi^*(t+1;\, t)\,[\,x(t) - x^*(t \mid t-1)\,] + u(t)$$
$$= \Phi^*(t+1;\, t)\tilde{x}(t \mid t-1) + u(t) \qquad (2.21)\ \{23, [\ 3]\}$$

The $E[u(t)] = 0$, thus equation (2.21) gives

$$E[\,\tilde{x}(t+1 \mid t)\,] = \Phi^*(t+1;\, t)\, E[\,\tilde{x}(t \mid t-1)]. \qquad (2.22)$$

Now, a sequence of steps utilizing the dynamic system within the right hand side of equation (2.22) gives $E[\tilde{x}(t+1 \mid t)]$ as a product of multiple transition matrices and $E[\,\hat{x}(t_o)]$. The expected value of the error in the original estimate, $E[\,\hat{x}(t_o)\,]$, is assumed to be zero, consequently,

$$E[\tilde{x}(t+1 \mid t)] = 0. \qquad (2.23)$$

Equations (2.21) and (2.23) yield a recursive relationship for the covariance matrix $P^*(t)$ of the error $\tilde{x}(t \mid t-1)$.

$$P^*(t+1) = E[\tilde{x}(t+1 \mid t)\tilde{x}'(t+1 \mid t)]$$
$$= \Phi^*(t+1; t)\, E[\tilde{x}(t \mid t-1)\tilde{x}'(t \mid t-1)]\, \Phi^{*\prime}(t+1; t) + Q(t)$$
$$= \Phi^*(t+1; t) P^*(t)\, \Phi^{*\prime}(t+1; t) + Q(t) \quad (2.24)\ \{24, [3]\}$$

where $Q(t) = E[u(t)\, u'(t)]$.

The remainder of the derivation requires an explicit formula for $\Delta^*(t)$.

Since $\tilde{x}(t+1 \mid t)$ is orthogonal to $\tilde{y}(t \mid t-1)$, it follows that

$$E[\tilde{x}(t+1 \mid t)\tilde{y}'(t \mid t-1)] = 0.$$

Expressing $\tilde{x}(t+1 \mid t)$ as

$$\tilde{x}(t+1 \mid t) = x(t+1) - \hat{E}[x(t+1) \mid \mathcal{Y}(t-1)] - \hat{E}[x(t+1) \mid \mathcal{Z}(t)] \quad (2.25)$$

and employing equation (2.16)

$$0 = E[(x(t+1) - \hat{E}[x(t+1) \mid \mathcal{Y}(t-1)] - \Delta^*(t)\tilde{y}(t \mid t-1))\tilde{y}'(t \mid t-1)]$$
$$= E[x(t+1)\tilde{y}'(t \mid t-1) - E[\hat{E}[x(t+1) \mid \mathcal{Y}(t-1)]\tilde{y}'(t \mid t-1)]$$
$$- \Delta^*(t)\, E[\tilde{y}(t \mid t-1)\, \tilde{y}'(t \mid t-1)]. \quad (2.26)$$

Because $\hat{E}[x(t+1) \mid \mathcal{Y}(t-1)]$ and $\tilde{y}'(t \mid t-1)$ are orthogonal, it follows that in equation (2.26)

$$E[\hat{E}[x(t+1) \mid \mathcal{Y}(t-1)]\tilde{y}'(t \mid t-1)] = 0.$$

The orthogonal projection of $x(t+1)$ onto $\mathcal{Y}(t-1)$ is denoted by $\bar{x}(t+1 \mid t-1)$, which is equal to $\hat{E}[x(t+1) \mid \mathcal{Y}(t-1)]$. Since $\bar{x}(t+1 \mid t-1)$ is in the linear manifold $\mathcal{Y}(t-1)$; it is orthogonal to $\mathcal{Z}(t)$. The error in the projection of $x(t+1)$

onto $\mathcal{Y}(t-1)$ is denoted by $\tilde{x}(t+1 \mid t-1)$. See Figure 4.

Thus, $x(t+1) = \bar{x}(t+1 \mid t-1) + \tilde{x}(t+1 \mid t-1)$.

From (2.26) and (2.17)

$$0 = E[(\bar{x}(t+1 \mid t-1) + \tilde{x}(t+1 \mid t-1))\tilde{y}'(t \mid t-1)]$$
$$-\Delta^*(t)\, E[M(t)\tilde{x}(t \mid t-1)(M(t)\tilde{x}(t \mid t-1))'].$$

Since $\bar{x}(t+1 \mid t-1)$ is orthogonal to $\mathcal{Z}(t)$ and by equations (2.1) and (2.2), and the definition of $P^*(t)$,

$$0 = E[\tilde{x}(t+1 \mid t-1)\tilde{y}'(t \mid t-1)] - \Delta^*(t)M(t)P^*(t)M'(t)$$
$$= E[\{\Phi(t+1; t)\,\tilde{x}(t \mid t-1) + u(t)\}\tilde{x}'(t \mid t-1)M'(t)$$
$$-\Delta^*(t)M(t)P^*(t)M'(t).$$
$$= \Phi(t+1; t)\,P^*(t)M'(t) - \Delta^*(t)M(t)\,P^*(t)M'(t).$$

The explicit formula for $\Delta^*(t)$ is

$$\Delta^*(t) = \Phi(t+1; t)P^*(t)M'(t)[M(t)P^*(t)M'(t)]^{-1}. \quad (2.27)\ \{25, [3]\}$$

The inverse will exist since $P^*(t)$ is positive definite and $M(t)$ is a $p \times n$ ($p \leq n$) matrix of rank p.

Thus, the solution given in Section 2 A is now verified.

3. Example

At this time an example will be presented to illustrate the concepts previously discussed. A dynamic system will be constructed in such a manner that statistical filtering techniques could be immediately applied. The model with additive noise in the observations is employed since it is more applicable to trajectory analysis. Approximation techniques in the transition matrix are introduced to provide an avenue for investigating an application in which the hypotheses of the theory are not completely satisfied.

The basic mathematical model is that of a particle in force free flight on a parabolic curve in the x-y phase plane. With this model, the statistical filter is used to estimate the position and velocity of the particle (i.e., the state of the dynamic system).

Whenever possible the notation of Section 2 will be used. The following standard notation will be needed;

$$\dot{x} = dx/dt,$$

$$\dot{y} = dy/dt,$$

$$\ddot{x} = d^2x/dt^2,$$

$$\ddot{y} = d^2y/dt^2,$$

$$y^{(n)} = d^ny/dt^n. \qquad n = 1, 2, 3, \ldots$$

The basic parametric equations which describe the motion of the particle are chosen to be:

$$\begin{aligned} x(t) &= 1 + at, \\ y(t) &= 1 + b\sqrt{t}, \end{aligned} \qquad t>0 \qquad (3.1)$$

where a and b may be considered as orbital parameters and the

independent variable t is considered as time. From equation (3.1),

$$\dot{x} = a,$$

$$\dot{y} = b/(2\sqrt{t}),$$

$$\ddot{x} = 0, \qquad (3.2)$$

$$\ddot{y} = -b/(4t^{3/2}) = -\dot{y}/(2t).$$

Choosing the components of the system to be the quantities $x_1(t) = x(t)$, $x_2(t) = y(t)$, $x_3(t) = \dot{x}(t)$, and $x_4(t) = \dot{y}(t)$; equations (3.1) and (3.2) produce a force free linear, time variant dynamic system

$$\dot{x}(t) = F(t)\, x(t) \qquad (3.3)$$

where

$$x(t) = (x_1(t), x_2(t), x_3(t), x_4(t))' = (x(t), y(t), \dot{x}(t), \dot{y}(t))',$$

$$\dot{x}(t) = (\dot{x}_1(t), \dot{x}_2(t), \dot{x}_3(t), \dot{x}_4(t))' = (\dot{x}(t), \dot{y}(t), \ddot{x}(t), \ddot{y}(t))',$$

and

$$F(t) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1/(2t) \end{pmatrix}. \qquad (3.4)$$

The next problem in working with a dynamic system is to determine the state transition matrix, that is the matrix, $\Phi(t; t_o)$, that relates the state vector at time t_o to the state vector at time t. More precisely one needs to determine a matrix $\Phi(t; t_o)$ such that $x(t) = \Phi(t; t_o)\, x(t_o)$.

This problem is fundamental in the theory of dynamic systems

and can generally be solved in several ways, some more appropriate than others. The initial step in determining the transition matrix is to solve the system of differential equations given in equation (3.3). In many cases the solution must be obtained by numerical methods. In this example, an analytic solution is easily obtained.

The general solution of equation (3.3) is

$$x(t) = A(t)c, \text{ for all } t > 0 \qquad (3.5)$$

where

$$A(t) = \begin{pmatrix} t & 1 & 0 & 0 \\ 0 & 0 & 2t^{\frac{1}{2}} & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & t^{-\frac{1}{2}} & 0 \end{pmatrix} \qquad (3.6)$$

and $c' = (c_1, c_2, c_3, c_4)$, a vector of arbitrary constants.

Given an initial condition $x(t_o)$ for the state vector, the technique is to solve for the arbitrary constant c and thus, eliminate c in equation (3.5).

Since A(t) is non-singular for $t > 0$, the inverse of A(t) exists and is

$$A^{-1}(t) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & -t & 0 \\ 0 & 0 & 0 & t^{\frac{1}{2}} \\ 0 & 1 & 0 & -2t \end{pmatrix} . \qquad (3.7)$$

In equation (3.5) allow t to be equal to t_o and multiply on the left by $A^{-1}(t_o)$ to obtain

$$c = A^{-1}(t_o)x(t_o). \qquad (3.8)$$

Substituting equation (3.8) into equation (3.5) gives

$$x(t) = A(t)\, A^{-1}(t_o)x(t_o). \qquad (3.9)$$

Equation (3.9) shows that $A(t)A^{-1}(t_o)$ is a transition matrix.

Let

$$\Phi(t;\, t_o) = A(t)A^{-1}(t_o)$$

$$= \begin{pmatrix} 1 & 0 & t-t_o & 0 \\ 0 & 1 & 0 & 2(t\, t_o)^{\frac{1}{2}}-2t_o \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & (t_o/t)^{\frac{1}{2}} \end{pmatrix} . \qquad (3.10)$$

Thus, $x(t) = \Phi(t;\, t_o)x(t_o)$. (3.11)

At this stage the transition matrix given in equation (3.11) is limited to relating the state vector at t_o to the state vector at time t. In using the transition matrix it is desirable to have $\Phi(t_2;\, t_1)$, $t_2 \geq t_1 > 0$, that is, a matrix satisfying the relationship

$$x(t_2) = \Phi(t_2;\, t_1)x(t_1).$$

The transition matrix $\Phi(t;\, t_o)$ possesses three properties:

Property (1)

$$\Phi^{-1}(t;\, t_o) = \Phi(t_o;\, t) \qquad (3.12)$$

This is easily derived: Since

$$\Phi(t;\, t_o) = A(t)\, A^{-1}(t_o),$$

the inverse of $\Phi(t;\, t_o)$ exists. Thus,

$$\Phi^{-1}(t;\, t_o) = (A(t)A^{-1}(t_o))^{-1}$$

$$= A(t_o)A^{-1}(t)$$

$$= \Phi(t_o;\, t).$$

The expression for $\Phi(t_2;\, t_1)$, is derived before proceeding

with the remainding two properties.

By equation (3.11)

$$x(t_2) = \Phi(t_2; t_o)\, x(t_o) \tag{3.13}$$

and $x(t_1) = \Phi(t_1; t_o)x(t_o)$.

Since $\Phi^{-1}(t; t_o)$ exists

$$x(t_o) = \Phi^{-1}(t_1; t_o)x(t_1). \tag{3.14}$$

Substituting (3.14) into (3.13) gives

$$\begin{aligned} x(t_2) &= \Phi(t_2; t_o)\Phi^{-1}(t_1; t_o)\, x(t_1) \\ &= \Phi(t_2; t_o)\Phi(t_o; t_1)x(t_1). \end{aligned}$$

Furthermore,

$$\begin{aligned} \Phi(t_2; t_1) &= \Phi(t_2; t_o)\Phi(t_o; t_1) \\ &= A(t_2)A^{-1}(t_o)\, A(t_o)\, A^{-1}(t_1) \\ &= A(t_2)\, A^{-1}(t_1) \\ &= \begin{pmatrix} 1 & 0 & t_2-t_1 & 0 \\ 0 & 1 & 0 & 2(t_1t_2)^{\frac{1}{2}} - 2t_1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & (t_1/t_2)^{\frac{1}{2}} \end{pmatrix}. \end{aligned} \tag{3.15}$$

Property (2)

$$\Phi(t_2; t_1)\Phi(t_1; t_o) = \Phi(t_2; t_o),$$

since

$$\begin{aligned} \Phi(t_2; t_1)\Phi(t_1; t_o) &= A(t_2)A^{-1}(t_1)A(t_1)A^{-1}(t_o) \\ &= A(t_2)A^{-1}(t_o) \\ &= \Phi(t_2; t_o). \end{aligned}$$

Property (3)

$$\Phi(t_o; t_o) = I,$$

since

$$\Phi(t_o;\ t_o) = A(t_o)\ A^{-1}(t_o)$$
$$= I,$$

where I is the identity matrix.

Reviewing, the particle in force free fight is represented by the set of differential equations given in equation (3.3), where F(t) is given in equation (3.4). The state transition matrix for any two times t_2 and t_1 is given in equation (3.15).

Now to further study the behavior of this dynamic system, an approximate transition matrix is developed. In physical problems approximations of the equations of motion are employed, thus this approximation illustrates this means of analysis.

The two main tools used in this development are Taylor series expansions and a recursive relation between $y^{(n)}(t)$ and $\dot{y}(t)$.

Notice in equation (3.2) $\ddot{x}(t) = 0$ and consequently, all the higher derivatives of x(t) with respect to the independent variable t will be zero. This is not true for y(t).

In equation (3.2)

$$y^{(2)}(t) = -y^{(1)}(t)/(2t)$$
$$= (-\tfrac{1}{2})^{2-1}(1)t^{-(2-1)}y^{(1)}(t). \qquad (3.16)$$

It follows from equation (3.2) that

$$y^{(3)}(t) = (-3/2)\ (-\tfrac{1}{4})t^{-5/2}$$
$$= (-\tfrac{1}{2})^{3-1}(3)\ (1)\ t^{-(3-1)}y^{(1)}(t).$$

By mathematical induction, it follows that

$$y^{(n)}(t) = (-\tfrac{1}{2})^{n-1}(2n-3)\ (2n-5)\ \ldots\ (3)\ (1)t^{-(n-1)}y^{(1)}(t). \qquad (3.17)$$

Equation (3.17) represents a recursive formula for the n^{th} derivative of y(t) with respect to the independent variable in terms of

$y^{(1)}(t)$. This formula will be used in the Taylor expansion of y(t).

One form of the Taylor series expansion is:

$$f(x + h) = \sum_{n=0}^{\infty} h^n f^{(n)}(x)/n! \quad .$$

Since y(t) satisfies the hypothesis of Taylor's theorem, for $h = t_2 - t_1$

$$y(t_2) = y(t_1 + h)$$

$$= \sum_{n=0}^{\infty} (t_2 - t_1)^n y^{(n)}(t_1)/n!. \qquad (3.18)$$

Substituting equation (3.17) into equation (3.18) gives

$$y(t_2) = y(t_1) + \sum_{n=1}^{\infty} (t_2-t_1)^n (-\tfrac{1}{2})^{n-1}(2n-3)(2n-5)\ldots(3)(1)t_1^{-(n-1)}y^{(1)}(t_1)/n!. \qquad (3.19)$$

For convenience equation (3.19) will be written as

$$y(t_2) = y(t_1) + y^{(1)}(t_1)\, S(t_2; t_1), \qquad (3.20)$$

where

$$S(t_2; t_1) = t_1 \sum_{n=1}^{\infty} ((t_2-t_1)/t_1)^n (-\tfrac{1}{2})^{n-1}(2n-3)(2n-5)\ldots(3)(1)/n!. \qquad (3.21)$$

It is easy to verify by D'Alembert's ratio test that this series is absolutely convergent for $t_2 < 2t_1$ and $t_1 < t_2$.

Analogous to equation (3.20), the relationship in equation (3.17) gives a Taylor series expansion

$$\dot{y}(t_2) = y(t_1)R(t_2; t_1) \qquad (3.22)$$

where

$$R(t_2; t_1) = \sum_{n=0}^{\infty} ((t_2-t_1)/t_1)^n (-\tfrac{1}{2})^n (2n-1)(2n-3)\ldots(3)(1)/n!. \qquad (3.23)$$

$R(t_2; t_1)$ is absolutely convergent for $t_2 < 2t_1$ and $t_1 < t_2$.

Define for $k \geq 1$

$$S_k(t_2;\ t_1) = t_1 \sum_{n=1}^{k} ((t_2-t_1)/t_1)^n(-\tfrac{1}{2})^{n-1}(2n-3)(2n-5)\ldots(3)(1)/n! \tag{3.24}$$

and

$$R_k(t_2;\ t_1) = \sum_{n=0}^{k} ((t_2-t_1)/t_1)^n(-\tfrac{1}{2})^n(2n-1)(2n-3)\ldots(3)(1)/n! \quad . \tag{3.25}$$

One can approximate $y(t_2)$ to any degree of numerical accuracy by equations (3.18) and (3.24) with an appropriate choice of k. Likewise $\dot{y}(t_2)$ can be approximated as close as desired.
The approximate transition matrix is defined as

$$\Phi_k(t_2;\ t_1) = \begin{pmatrix} 1 & 0 & t_2-t_1 & 0 \\ 0 & 1 & 0 & S_k(t_2;\ t_1) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & R_k(t_2;\ t_1) \end{pmatrix} \quad . \tag{3.26}$$

It should be noted that the approximate Φ matrix $\Phi_k(t_2;\ t_1)$ does not have all of the properties of the exact transition matrix. For instance, in the approximate matrix,

$$\Phi_k(t_2;\ t_0) \neq \Phi_k(t_2;\ t_1)\Phi_k(t_1;\ t_0).$$

Consequently, some of the error in the estimation is due to the matrix $\Phi_k(t_2;\ t_1)$ and not the filtering procedure.

The dynamic system also includes the equation which connects the observable vector y(t) and the state vector x(t), i.e.,

$$y(t) = M(t)x(t) + u(t) \tag{3.27}$$

where, M(t) is the mapping matrix and u(t) is a vector of random "noise" or errors whose statistics are known.
$E[u(t)] = 0$ and $E[u(t)u(t)'] = E[\{y(t) - M(t)\ x(t)\}\{y(t)-M(t)x(t)\}'] = Q(t)$.

Computer studies are now being conducted to evaluate the results

obtained by using the approximate transition matrix in the statistical filtering scheme illustrated in Figure 1. Preliminary data indicates that under certain conditions the results are very good.

4. Conclusions and Future

In summary, a basic conclusion is that a Kalman type filter gives the "best" minimum variance estimator when the hypotheses of the theory are satisfied. Other statistical techniques will also yield a minimum variance estimator. But, a Kalman type filter routine is an updating process; consequently, it has certain advantageswhen considered in regard to computational requirements. The advantages are, in particular, smaller core storage needs and smaller matrix inversions.

In non-linear application problems, a Kalman type filter may be employed under the provision that the non-linear structure is linearized about a reference. This application is reasonable when the deviates from the reference to the actual trajectory that forms the mathematical model of the linear system are small. In some problems an updating routine to produce new references is needed to keep the deviates within the required bounds.

Future study is needed to investigate questions that arise in application techniques. These questions are mainly in regard to the linearization processes and computor calculations. In structure these may be classified as being in three areas: the first area is the effect of the linearization of a non-linear process; the second area is computational effects within calculations of the transition equation; and the last general area is the effect of the measurement functions, its linearization or deficiencies in the calculations.

The philosophy of this further study should be directed toward the effects of utilizing the mathematical theory in situations where the hypotheses of the structure are not completely satisfied.

The authors of this report feel that these questions can be investigated via the avenue of simple mathematical models. The goal of these investigations would be to gain insight into the effect of failing to satisfy particular hypotheses of the theory. In particular, these investigations could involve the following points: the updating schedule of the reference in the linearization of a non-linear problem, the propagation of errors, the overall effect of observation increments, and the requirements associated with non-linear measurement functions. For example, a task problem could be to structure a non-linear mathematical model--linearize the model and apply a Kalman type filter, then use a non-linear filter that has been approximated for computational purposes. This comparison could generate considerable insight. Also in some application problems the covariance matrix goes negative definite, an impossibility when all the hypotheses are satisfied. The model could also be structured in such a manner that a hint toward the reason for this dilemma could be found.

Abstracts of Selected Reports

The following are abstracts of references [1], [5], [10], and [12]. The abstracting of these articles was prescribed by the contracting agency as one of the objectives of the contract. In this section the equation numbers in braces refer to the corresponding equation in the paper being abstracted.

J P L Space Programs Summary No. 37-21, B. Dynamic Filtering, [1].

The author of this article attempts to clear-up some of the difficulties in reading and understanding R. E. Kalman's original paper [3] concerning a particular formulation to the solution of the "Wiener Problem." The level of abstraction of the article is reduced considerably compared to Kalman's paper.

By using varied mathematical assumptions, the author is able to obtain equivalent equations by a less abstract method. The "optimization" is based on the following theorem.

Given two vectors of random variables, say x and y, with zero mean vectors; $\Lambda_x = E[xx']$, $\Lambda_y = E[yy']$, and $\Lambda_{yx} = E[yx']$, the "best" linear estimate of y, say $\hat{y}$, given the value of x is

$$\hat{y} = \Lambda_{yx}\Lambda_x^{-1}x. \qquad \{1\}$$

"Best" means that in the class of all linear estimators (linear functions of the known vector x) $\hat{y}$ has the covariance matrix with minimum trace.

Letting $Q = E[(y - \hat{y})(y - \hat{y})']$, the "minimized" covariance matrix of $\hat{y}$ is given by

$$Q_m = \Lambda_y - \Lambda_{yx}\,\Lambda_x^{-1}\,\Lambda_{xy}. \qquad \{2\}$$

The author uses this theorem to solve the following problem. Let the dynamic equations of the mathematical model be expressed by

$$x(t_{n+1}) = \Phi(t_{n+1};\, t_n)\, x(t_n), \qquad \{3\}$$

$$y(t_n) = M(t_n)\, x(t_n) + u(t_n), \qquad \{4\}$$

where

$x(t_n)$ = state vector at time t_{n+1},

$M(t_n)$ = mapping matrix,

$\Phi(t_{n+1}; t_n)$ = transition matrix,

$y(t_n)$ = observables of the system,

$u(t_n)$ = random noise with zero mean vector and covariance matrix $Q(t_n)$.

Basically, the problem is, how does one obtain a "best" estimate of the error, at time t_{n+1}, that is, $\hat{x}(t_{n+1})$, when

$$\hat{x}(t_{n+1}) = x(t_{n+1}) - \bar{x}(t_{n+1}),$$

$$\bar{x}(t_{n+1}) = \Phi(t_{n+1}; t_n)\, x^*(t_n),$$

and $x^*(t_n)$ is the best estimate of $x(t_n)$ based on the observations $y(t_1), \ldots, y(t_{n-1})$.

The answer to the above problem is found by using equation {1}. That is, one needs to find corresponding terms in the right hand side of the equation {1}. The author does this using only elementary statistics and matrices.

The development has two basic steps. The first step is the prediction of the state of the system x_{n+1} at the next time step t_{n+1} based on a best estimate of $x^*(t_n)$ of the current state, using only the observations $y(t_1), \ldots, y(t_{n-1})$.

The second step is correcting the predicted state, after taking the measurements at time t_n. The "updating" procedure in step two uses the stated theorem.

The remainder of the article is mainly concerned with developing recursive formulas for updating the covariance matrix of the

estimate $\hat{x}_{n+1}$.

The notation in this article is essentially that of Kalman's original paper [3].

An Introduction to Midcourse Navigation-Guidance, Autonetics [5].

This report is written with respect to the on board system philosophy. The midcourse navigation guidance problem is composed of; the trajectory determination problem, the attitude determination problem, the prediction problem, and the trajectory control problem. These problems are well defined in the report. The Kalman filter is used for the trajectory determination problem and as a starting point for trajectory control.

The Kalman filter is discussed mainly in chapters two and three. The notation is essentially that of Kalman's. A heuristic derivation of Kalman's filter is presented. There is also a development (not found in the other reports abstracted) of asymptotic cases. Asymptotic cases arise when the predictions are poor and the observations good; and when the predictions are good and the observations bad.

The author assumes (i) the space craft is sufficiently close to the reference trajectory (this means linear perturbation methods may be used), and (ii) the derivation state is measured relative to the reference trajectory.

The mathematical model utilized is

$$y_n = M_n x_n , \qquad \{2.1\}$$

$$x_{n+1} = \Phi_{n+1,\,n}\, x_n + \Delta_{n+1,\,n}\, u_n , \qquad \{2.2\}$$

where y_n = observation residuals vector for the n^{th} navigation-guidance cycle,

x_n = position-velocity deviation state vector at the beginning

of n^{th} cycle,

M_n = matrix relating current position-velocity deviation to observation residuals,

$\Phi_{n+1,n}$ = state transition matrix,

u_n = control function (velocity correction),

$\Delta_{n+1,n}$ = matrix relating control function at n^{th} step to control function at $n+1^{st}$ step.

It should be noted that x_n is a deviation state vector and that y_n is a residual vector. Equation {2.1} is used for the trajectory determination problem, i.e., estimate x_n given y_n.

The heuristic development of the Kalman filter assumes

$$\hat{y}_n - \hat{y}'_n = \omega^*_n (\tilde{y}_n - \hat{y}'_n) \qquad \{3.3\}$$

where ω^*_n is a weighing filter. This equation means that the best estimate of the n^{th} observation residuals can be equated to the best predicted observation residual $\hat{y}'_n$, plus an optimizing filter matrix ω^*_n times the difference between the measured and predicted observation residuals, $(\tilde{y}_n - \hat{y}'_n)$. That is to say, the best estimate of the difference between the observation residuals is equal to the optimal weighing function ω^*_n times the difference between the measured and predicted values.

The objective is to choose $\hat{y}_n$ so that its difference from the actual observation residual, y_n, is minimized according to some loss criterion. The loss criterion used to compute ω^*_n is the mean square deviation. For convenience, Autonetics minimizes the trace of the covariance of $(\hat{y}_n - \hat{y}^A)$. In minimizing the trace then, heuristically, the covariances are reduced. Minimizing the trace

leads to results in terms of covariance matrices.

The optimal filter is found to be

$$\omega_n = C(\Delta\hat{x}_n') M_n' [M_n C(\Delta \hat{x}_n') M_n' + C (\Delta\tilde{y}_n)]^{-1}. \qquad \{3.22\}$$

An alternate form can be expressed as

$$\omega_n = [C(\Delta\hat{x}_n')^{-1} + M_n' C(\Delta\tilde{y}_n)^{-1} M_n]^{-1} M_n' C(\Delta\tilde{y}_n)^{-1} \qquad \{3.23\}$$

Conditions are stated for equivalence of {3.22} and {3.23} in appendix A.

The optimal estimate trajectory deviation equation can be expressed as a single recurrence relation,

$$\hat{x}_n = (I - \omega_n M_n) (\Phi_{n,\, n-1} + \Delta_{n,\, n-1}\, a_{n-1}') \hat{x}_{n-1} + \omega_n \tilde{y}_n. \qquad \{3.21\}$$

This result anticipates the nature of the optimal control u_{n-1} at the beginning of the $(n-1)^{th}$ cycle.

An alternate form of {3.21} is given by

$$\hat{x}_n = [C(\Delta\hat{x}_n')^{-1} + M_n' C(\Delta\tilde{y}_n)^{-1} M_n]^{-1} [C(\Delta\hat{x}_n')^{-1} \hat{x}_n' + M_n' C(\Delta\tilde{y}_n)^{-1}\tilde{y}_n]. \qquad \{3.26\}$$

In this form, the two inputs to the optimal estimate computation are the predicted estimate x_n' weighted by the covariance of the prediction error plus the observation data y_n weighted by the covariance of its errors.

Asymptotic cases are discussed and put in tabular form. If the covariance of the errors in the prediction is large, i.e., the prediction is poor, then $C(\Delta\hat{x}_n') >> C(\Delta\tilde{y}_n)$. Symbolically the inequality means diagonal elements are large. If the prediction is reliable but the observation errors are large then $C(\Delta\tilde{y}_n) >> C(\Delta\hat{x}_n')$.

Autonetics not only discuss the asymptotic cases but also

emphasize the differences between the even-or-over determined set of observations and the under-determined case. Basically, the decision depends upon the quality of the current set of observation residuals. From the computational point of view, the under-determined case leads to inversions of matrices of orders less than 6.

There are many typographical mistakes in the Autonetics report and caution should be used, especially in the asymptotic cases.

Definitions

$C(\Delta\hat{x}'_n) = E[(\hat{x}'_n - x^A_n)(\hat{x}'_n - x^A_n)']$ = covariance matrix on dispersion of predicted estimate,

$C(\Delta\tilde{y}_n) = E[(\tilde{y}_n - y^A_n)(\tilde{y}_n - y^A_n)']$ = covariance matrix on observation errors,

a'_{n-1} = control filter and is defined by the equation $\hat{u}_{n-1} = a'_{n-1}\hat{x}_{n-1}$.

The Application of State Space Methods to Navigation Problem, Technical Report No. 4, Philco Western Development Laboratories, [10].

The topic of main concern is determining the state of a system; therefore, the material on pages 36 through 56 of the above mentioned technical report will be considered. This material is directly concerned with the derivation and application of the Kalman statistical filter.

This section is prefaced with the following remarks. The material immediately preceding this section of the report introduced the standard statistical terms such as expected value, density function, and multivariate normal (Gaussian) distribution. The covariance matrix, P, introduced on page 31 in defining the multivariate normal distribution must be a positive definite matrix. The fact that P is positive definite will play an important role in the development of the Kalman filter.

Because of insufficient observable quantities and the inaccuracy in measurements, it is necessary to "smooth" the observed data, that is "filter" out the "noise" in the observations. This means that the data must be processed in such a way so as to remove or eliminate the effect of the random errors. R. E. Kalman [3] derived a method of solving this problem for linear systems and others [2], [11] were able to develop a technique to apply Kalman's solution to non-linear

systems.

This technique needs an initial estimate of the state of the dynamic system and a covariance matrix for the error in this estimate. The report presents one method of obtaining the initial estimate, the method of "least squares fit to a polynomial."

To derive the least squares equations, the "gradient" was introduced.

Suppose n observations of a variable X(t) are taken (i.e., $X(t_1)$, $X(t_2)$, ..., $X(t_n)$) and X(t) is to be represented as a polynomial of degree $m < n$ in t. That is

$$X(t_1) = y_0 + y_1 t_1 + y_2 t_1^2 + \ldots + y_m t_1^m,$$

$$X(t_2) = y_0 + y_1 t_2 + y_2 t_2^2 + \ldots + y_m t_2^m,$$

$$\vdots$$

$$X(t_n) = y_0 + y_1 t_n + y_2 t_n^2 + \ldots + y_m t_n^m,$$

or representing these equations as a matrix equation

$$X = Ay \qquad (1)$$

where

$$X = \begin{pmatrix} X(t_1) \\ X(t_2) \\ \vdots \\ X(t_n) \end{pmatrix}, \quad A = \begin{pmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^m \\ 1 & t_2 & t_2^2 & \cdots & t_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^m \end{pmatrix}, \quad y = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

But the y vector is unknown and must be determined. The criteria of least squares requires one to choose constant $y_1, y_2, \ldots, y_m$, such that

$$L = \sum_{k=1}^{n} (X(t_k) - (y_o + y_1 t_k + \ldots + y_m t_k^m))^2 \qquad (2)$$

is a minimum. Equation (2) represents the sum of the squares of the vertical deviations from the fitted line. In terms of equation (1), equation (2) is

$$L = (X - Ay)' (X - Ay). \qquad (3)\{77\}$$

By using the gradient technique to minimize equation (3) with respect to the choice of the vector y, one obtains the the least squares choice for y, denote by $\hat{y}$ as,

$$\hat{y} = (A'A)^{-1}A'X. \qquad (4)\{81\}$$

Consequently one can now estimate X(t) at some other time t and the accuracy of this estimate will depend on the accuracy of the original observations,

$$X(t_1), \ldots, X(t_n).$$

If the original observations have a known covariance matrix

$$P_x = E[(X - E[X]) (X - E[X])']$$

then the covariance matrix of $\hat{y}$ is

$$P_{\hat{y}} = (A' A)^{-1} A' P_x A(A' A)^{-1}. \qquad (5)\ \{82\}$$

Then from equations (4) and (5) the initial estimate and covariance matrix required by the Kalman filtering technique can be obtained.

Suppose one has a linear system and an estimate of the

state vector. The Kalman filter is basically a scheme to update or improve the estimate of the state vector after a set of new observations are taken.

Schmidt states the problem as:

"Given $\hat{x}(t_1)$ = estimate of x,

$P(t_1) = E(x - \hat{x})\,(x - \hat{x})'$ = covariance matrix of the error in the estimate,

$y(t_1) = Hx(t_1) + q(t_1)$,

$q(t_1)$ = random error in measurement of $y(t_1)$,

$E(q(t_1)) = 0$,

$E(q(t_1)q'(t_1)) = Q$.

Find: a new estimate $\hat{x}_n(t)$ of $x(t_1)$ such that

$L = E(x - \hat{x}_n)'\,(x - \hat{x}_n)$ is minimized." {85}

This formulation of the problem seems incomplete. In no way does it bring in the fact that one is trying to update the present estimate by using the information in a new set of observations. The notation is confusing. For example, the new estimate at time t_1, is denoted by "$\hat{x}_n(t)$."

This article gives two derivations of the Kalman filter, one assuming that the random variables are normally (Gaussian) distributed and the second requires that the estimate be a linear function of the observations (a linear filter).

The notation in derivation one is misleading.

For example,

$$L = \int (x - \hat{x}_n)'\,(x - \hat{x}_n)\; p(x \mid y, \hat{x})\; dx\,, \qquad \{86\}$$

indicates that the state vector is a random variable with conditional probability density function $p(x \mid y, \hat{x})$. The random variable is actually the <u>error</u> $x - \hat{x}$.

The "results" are given in equations (6) and (7), that is;

$$\hat{x}_n = \hat{x} + PH' (HPH' + Q)^{-1} (y - \hat{y}), \qquad (6)\ \{95\}$$

where $\hat{y} = H\hat{x}$ and

$$P_n = P - PH' (HPH' + Q)^{-1} HP. \qquad (7)\ \{99\}$$

Equation (7) is the method for finding the covariance matrix of the new estimate.

In the second derivation the development is straight forward and requires mostly elementary statistics and matrix theory.

The remainder of the material considered in this report is an example which illustrate the use of the filtering equations, (6) and (7).

Capabilities of MSFN for Apollo Guidance and Navigation, Sections 3.0, 3.1, and Appendixes A, B., Bissett-Berman Corp., [12].

The authors of these particular sections of the report direct their attention toward a discussion and description of the analytical techniques that are employed in statistical filtering. They select as the avenue of approach the explanation and critique of simple mathematical examples. The first example is the estimating of the parameters of an "orbit" $y = a + bt$, and the second example is the estimation of the parameter a in a simple sine curve $y = \sin at$.

The authors assume a variable $y(t)$, dependent of time via a known relationship that involves unknown parameters. The parameters are to be estimated using a sequence of $n + 1$ measurements, $y_m(t)$ $t = 0, 1, \ldots, n$, that contain errors or "noise." The errors are assumed to be uncorrelated, have zero mean, and form a stationary Gaussian process with standard deviation σ.

In the linear problem, $y = a + bt$, a straight forward maximum likelihood technique is used to obtain the best estimates for a and b as a function of the measured values. The variances and covariance of these estimates are also derived.

In a situation with two parameters, a and b, the maximum likelihood values, $\underline{a}$ and $\underline{b}$, are those that maximize the likelihood function $p(a, b)$. Here

$$p(a, b) = \frac{1}{(2\pi)^{\frac{1}{2}n + \frac{1}{2}n+1}} \exp(-\tfrac{1}{2}\ \mathcal{I}),$$

in which

$$\mathcal{I} = \sum_{t=0}^{n} [y_m(t) - y(t)]^2/\sigma^2.$$

In the example, $y = a + bt$, the estimates $\hat{a}$ and $\hat{b}$ of the parameters a and b are found by taking partial derivatives of the $\mathcal{I}$ function and determining the critical points. The estimator equations are thus derived. The relationships shown generate estimates with errors that have zero mean and simple variances and covariance expressions. The mathematics is straight forward.

The sine example is selected to illustrate the problems of linear estimates employed in non-linear situations. Here a nominal value a_o is selected and an improved value, a_1, is desired. The best linear estimate is calculated and it is demonstrated in appendix B that this value is not the maximum likelihood value. In fact its accuracy is dependent on the value of the original estimate and extensive data input will not assure an approach to the desired value. The authors conclude that successive iterations are needed to approach the maximum likelihood value. Then the statement is made that J P L and Goddard have found that three iterations have been sufficient to bring the calculated values close enough to the maximum likelihood values. The analysis of this example is carried further to illustrate that the statistical uncer-

taintities of the maximum likelihood estimates can be obtained without computing the maximum likelihood values.

In the second example there is one parameter, a, and the likelihood function is then

$$p(a) = \frac{1}{(2\pi)^{(n+1)/2}\ \sigma^{n+1}} \exp(-\tfrac{1}{2}\ \mathcal{I})$$

where $\mathcal{I} = \sum_{t=0}^{m} [y_m(t) - y(t)]^2 / \sigma^2$.

Again the derivative is calculated to arrive at the critical point that will minimize $\mathcal{I}$ and consequently maximize p(a). This generates an equation that is difficult to solve and a linearization about an initial value a_o of the parameter a is employed. The linearized equation to be solved for the linear best estimate is arrived at by substituting in the lower order terms of the Taylor series expansion of the functions. This estimate is used in the discussions of the pitfalls of the best linear estimates.

The authors conclude that a best linear estimate scheme does not give the desired maximum likelihood value. In particular they conclude that a good original estimate is needed in order that the estimate will be reasonably accurate and a linear best estimate in a non-linear application is limited to those cases in which a good a prior knowledge of the parameters is available.

The body of appendix A is a derivation of the best linear estimates arrived at via a Kalman approach and a linear least

square analysis. Then a demonstration is given that the two results are equivalent under a hypothesis that the state variables are independent of time.

In appendix B the authors demonstrate that the linear best estimate of the parameter, a, in the sine example is not the correct value. The difference in the answers arises from the fact that the actual and reference "trajectories" diverge as time increases.

In the derivation of the Kalman-Schmidt filter a fixed reference trajectory is assumed and the filter is derived for deviates about this reference. The notation is essentially standard with a few variations employed. That is, x_n is the state vector of deviates from the reference trajectory, y_n is the state vector of the deviates of the observables, $\Phi_{n+1,n}$ is the state transition matrix that updates the x_n vector to x_{n+1}, M_n relates y_n to x_n, Q_n is the covariance of the noise, and P_n^* is the covariance matrix of the errors in the estimated state x_n^* and the state x_n. The derivation of the filter equations is given via algebraic manipulations and application of a "fundamental theorem." The fundamental theorem is a variation of the widely used Gauss-Markov theorem that expresses the minimum variance linear estimate, $\hat{\beta}$, of β given α as

$$\hat{\beta} = \Lambda_{\beta\alpha} \Lambda_{\alpha}^{-1} \alpha$$

where $\Lambda_{\beta\alpha}$ and Λ_{α} are covariance matrices associated with the random variable vectors α and β.

The basic technical idea involved is the linearization of a non-linear problem and thus being able to apply linear estimation theory. The authors are critical of this type of application and they devote most of their discussion to an examination of its weaknesses. Their points should be well taken, but the general hypotheses of this situation should also be remembered. Any application of this type will always generate unsafe conclusions if the hypotheses are violated.

First of all, Kalman's original derivation is built of the hypothesis of a linear dynamic model. The specific conclusion that Kalman gives is a minimum mean square error estimate for the system as a function of the sequence of observables. The second consideration is that any linearization of a non-linear system must be very cognizant of the linearization process. The general approach is to select a reference relationship and investigate the deviates of the actuals from the reference. In case the deviates are within bounds, the deviates can be expressed as a linear dynamic system with the first order terms of the Taylor series expansions.

In general, one would have to assume that after some period of time the actual and reference situations would vary away from each other. Consequently the first order term approximations of the deviates would no longer be accurate. Thus it is not the linear filter that is in error, it is the fact that the deviates are too large to accurately apply only to the

first order terms of the series. At this stage one must have some sort of routine for selecting a new nominal. If not, one should expect and will probably get inaccuracies.

Bibliography

1. Solloway, C. B.: Dynamic Filtering. Jet Propulsion Laboratory Space Programs Summary No. 37-21, Pasadena, California, March, 1963. pp. 7-10.

2. Smith, Gerald L., Schmidt, Stanley F., and McGee, Leonard A.: Application of Statistical Filter Theory to the Optimal Estimation of Position and Velocity on Board a Circumlunar Vehicle. NASA TR R-135, 1962.

3. Kalman, R. E.: A New Approach to Linear Filtering and Prediction Problems. Trans. ASME, Ser. D, Jour. Basic Engr., vol. 82, no. 1, Mar. 1960, pp. 35-45.

4. Liebelt, P. B., and Bergeson, J. E.: The Generalized Least Square and Wiener Theories With Applications to Trajectory Prediction. Boeing Document No. D2-90167, May, 1962.

5. Kochi, K. C.: An Introduction to Midcourse Navigation Guidance. Autonetics No. T-230/32-3.

6. Wiener, N.: The Extrapolation, Interpolation and Smoothing of Stationary Time Series. John Wiley & Sons, Inc., New York, 1949.

7. Kalman. R. E.: Mathematical Description of Dynamical Systems. Jour. Soc. Indust. Appl. Math. Ser. A Control vol. 1, no. 2, 1963. pp. 152-192.

8. Joyce, L.: A Dynamic Filtering Problem. Jet Propulsion Laboratory Space Programs Summary No. 37-28, vol. IV, pp. 8-10.

9. Doob, J. L.: Stochastic Processes. John Wiley & Sons, Inc. New York, 1953.

10. Schmidt, Stanley F.: The Application of State Space Methods to Navigation Problems. Philco Western Development Laboratories Technical Report No. 4. July 1964.

11. McLean, John D., Schmidt, Stanley F, McGee, L. A.: Optimal Filtering and Linear Prediction Applied to a Midcourse Navigation System for the Circumlunar Mission. NASA TN D-1208, March 1962.

12. Bissett-Berman Corporation: Capabilities of MSFN for Apollo Guidance and Navigation. March, 1964.